

A novel, web-based, parallel concordancer for use in the ESL/EFL classroom

Laurence Anthony* Kiyomi Chujo Kathryn Oghigian
 Waseda University Nihon University Waseda University

<http://www.antlab.sci.waseda.ac.jp/>
 anthony@waseda.jp



Faculty of Science and Engineering, Waseda University

Outline

- Background
 - Nihon University, College of Industrial Technology Communication Course
 - Software Needs
- Overview *WebConc-Bilingual*
 - System features
 - Google-like database architecture
 - Postings-merging algorithm
- Pedagogical Applications

2

Background (1)

- Nihon University, College of Industrial Technology Communication Course
 - Student Level
 - Student range TOEIC-IP 290-520 (≈TOEFL PBT 397-477)
 - Goals
 - Understand basic grammar and collocation usage
 - Develop basic vocabulary for communication
 - Approach
 - Data-driven Learning (DDL) using a parallel corpus of English/Japanese newspaper texts (Utiyama & Isahara, 2003)
 - Current Software
 - Paraconc (Barlow, 2004)



3

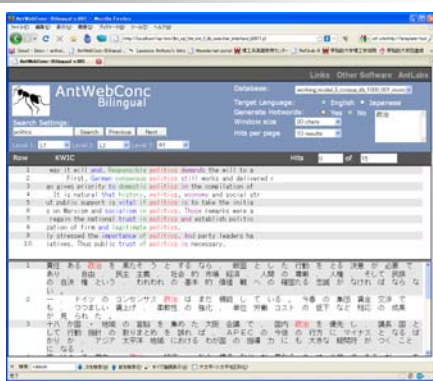
Background (2)

- Software Needs (based on student feedback)
 - Improved usability
 - designed for learners (not researchers)
 - quick startup (e.g., loading corpus, initializing settings, ...)
 - improved visualization (e.g., larger font size, fewer garbled characters, ...)
 - Improved access
 - web-based (for easy access anywhere, anytime, any OS)
 - freeware (for students with limited budgets)
 - Improved scalability and speed
 - handle larger corpora (in the future)
 - mimic the response time of traditional web search engines

4



WebConc -Bilingual-



5



WebConc -Bilingual-

- Usability
 - intuitive interface designed for learners (and teachers)
 - fully compliant with Unicode character classes
 - fewer garbled characters
 - no cumbersome token definition settings
 - powerful wildcard support (based on regular expressions)
- Access
 - web-based, Open Source
 - LAMP (Linux, Apache, MySQL, PHP) framework
 - fully compliant with HTML 4.01 and Section 508 accessibility
- Scalability and speed
 - Google-like database architecture

6

WebConc
-Bilingual-

TEXT(S) IN DATABASE

40. We support the phased strategy of cooperation between the Russian government and the IMF.

11. We welcome the progress made in the nuclear safety program agreed at the Munich Summit, including the establishment of the multilateral fund, in which we encourage broader participation.

36. All summit participants have shown solidarity in a critical situation by providing extensive food aid, credits and medical assistance.

11. International institutions have been central to our pursuit of stability, prosperity and equity for the past 50 years.

23. We confirm the validity of the international debt strategy.

POSTINGS LIST

...					
We	40:1	45:7	85:5,9	90:1,15	...
support	1:5,18	12:88	40:2	90:2,37	213:39 ...
the	13:1,8,53	40:3,9,13	11:3,7,13,18,21	...	
phased	55:58	40:3,9,13			
strategy	40:5	DocID, pos ₁ , pos ₂ , pos ₃ , ...			
of	16:4,68,215	40:6	82:263	...	
cooperation	1:10,38	18:4,28,88	40:7	185,4,16	...
between	19:8,44,137	40:8	185:3,17...		
Russian	31:1,22,64	40:10	311:46,141	...	
...					

The Anatomy of a Large-Scale Hypertextual Web Search Engine (Brin & Page, 1998)

7

WebConc
-Bilingual-

TEXT(S) IN DATABASE

40. We support the phased strategy of cooperation between the Russian government and the IMF.

11. We welcome the progress made in the nuclear safety program agreed at the Munich Summit, including the establishment of the multilateral fund, in which we encourage broader participation.

36. All summit participants have shown solidarity in a critical situation by providing extensive food aid, credits and medical assistance.

11. International institutions have been central to our pursuit of stability, prosperity and equity for the past 50 years.

23. We confirm the validity of the international debt strategy.

POSTINGS LIST

...					
We	40:1	45:7	85:5,9	90:1,15	...
support	1:5,18	12:88	40:2	90:2,37	213:39 ...
the	13:1,8,53	40:3,9,13	11:3,7,13,18,21	...	
phased	55:58,208	40:4	56:11,34	...	
strategy	40:5	130:2,32	5557:209	...	
of	16:4,68,215	40:6	82:263	...	
cooperation	1:10,38	18:4,28,88	40:7	185,4,16	...
between	19:8,44,137	40:8	185:3,17...		
Russian	31:1,22,64	40:10	311:46,141	...	
...					

The Anatomy of a Large-Scale Hypertextual Web Search Engine (Brin & Page, 1998)

8

WebConc
-Bilingual-

■ Postings Merge Algorithm (Manning et al., 2008: 42)

```

1  We support the phased strategy of cooperation between the Russian government and the IMF.
2  We welcome the progress made in the nuclear safety program agreed at the Munich Summit, including the establishment of the multilateral fund, in which we encourage broader participation.
3  All summit participants have shown solidarity in a critical situation by providing extensive food aid, credits and medical assistance.
4  International institutions have been central to our pursuit of stability, prosperity and equity for the past 50 years.
5  We confirm the validity of the international debt strategy.

```

Example

We
40:1 45:7 85:5,9 90:1,15 ...

support
1:5,18 12:88 40:2 90:2,37 213:39 ...

Hits (we support)
40:1-2 90:1-2

9

WebConc
-Bilingual-

10

AntWebConc Bilingual

Database: working_model_3_corpus_db_001_monospace

Target Language: English Japanese

Generate Hotwords: Yes No

Window size: 50 chars

Hits per page: 100 results

Level 1: R1 Level 2: R2 Level 3: R3

Row KWIC Hits 0 of 0

12

Pedagogical Applications

12

Syllabus (2009)

Week	Spring Semester		Fall Semester	
	Grammar	Vocabulary	Grammar	Vocabulary
1	Word classes	Business [1]	Structure of VP [1]	Business
2	derivations and inflections	Business [2]	Structure of VP [2]	Personnel [1]
3	Count & non-count nouns	Personnel [1]	Structure of VP [3]	Personnel [2]
4	Structure of NP [1]	Personnel [2]	Structure of VP [4]	Meetings
:	:	:	:	:
10	Structure of NP [7]	Daily Life	Structure of VP [10]	Transportation

Syllabus (2009)

Noun Phrases	Verb Phrases
NP: art + adj + n	VP: gerunds & infinitives
NP : followed by PP	VP: that clause
NP: followed by infinitive	VP: adverb location
NP: followed by -ed, -ing	VP: iv. + complement

14

A Japanese-English Parallel Corpus with ParaConc

Search **a * opportunity for** and find what word comes before **opportunity**.

限定詞	形容詞	名詞	名詞の直後
1 ()	()	opportunity	for Japan
2 ()	()	opportunity	for a review ...
3 ()	()	opportunity	for finally resolving ...
4 ()	()	opportunity	for the globalization ...
5 ()	()	opportunity	for the party
6 ()	()	opportunity	for the government
7 ()	()	opportunity	for us
8 ()	()	opportunity	for us

a tapering off it offers a good opportunity for Japan to reexamine its pro-
 reform stance will provide a good opportunity for a review of the police archi-
 current problem presented a good opportunity for finally resolving the racism
 y whom could be viewed as a good opportunity for the abolition of the year
 into liberal parties offers a good opportunity for the party to burge itself in
 visit. The meetings provide a good opportunity for the government to listen to
 i vote. The contest offers a good opportunity for us to think about who will i
 i Fresh Club on Monday a good opportunity for to learn whether to accept
 a praised the festival as a great opportunity for Japan and Britain to deepen
 tennis had a visit will see a great opportunity for Japanese and British leaders
 TE. At least had been given a new opportunity for self-government and interna-
 parian relations afforded a unique opportunity for radical Islam to enter the
 日本「一党独裁」を断絶する「憲法」論争の機会に、後述の通り「憲法」が守
 1の機会を、その見直しを、機会に、この機会に、この機会に、この機会に、
 三党の間の交渉、機会に、この機会に、この機会に、この機会に、この機会に、
 FUSAの間の交渉に、この機会に、この機会に、この機会に、この機会に、
 最も、この機会に、この機会に、この機会に、この機会に、この機会に、
 業の間に、この機会に、この機会に、この機会に、この機会に、この機会に、
 その機会に、この機会に、この機会に、この機会に、この機会に、この機会に、
 機会に、この機会に、この機会に、この機会に、この機会に、この機会に、
 。 機会に、この機会に、この機会に、この機会に、この機会に、この機会に、

15

Class Results (Spring 2009)

	DDL		Non-DDL
	DDL + Vocabulary Class 1 (TB 69%)	DDL + Vocabulary Class 2 (TB 56%)	Listening + Vocabulary (TB 54%)
Vocabulary test	31%**	37%**	36%**
Test to identify NP (easy words)	12%**	20%**	7%*
Test to identify NP (TOEIC words)	27%**	27%**	10%*
Producing NP test	19%**	15%**	3%
TOEIC-type NP test	9%*	2%	2%

TB: TOEIC Bridge

**statistically significant at the 1% level

* 5% level

16

Summary

- A fast, web-based, Google-like, parallel concordancer is being developed to allow students to carry out DDL exercises inside and outside the classroom
- The software will be applied in a Japanese university class where students improve their grammar skills using a Japanese-English parallel newspaper corpus
- **Acknowledgments**
This study is funded by a Grant-in-aid for Scientific Research (21320107) from the Japan Society for the Promotion of Science and Ministry of Education, Science, Sports and Culture.

17